

REMARKS

This Amendment is in response to the Office Action of June 26, 2008 in which all of the pending claims 1-6 and 9-20 were rejected.

State of the Art

Fortier et al. (US 6,584,179 B1) describes in general a method of improving the utility of speech recognition of spoken words and an apparatus for improving thereof. In general, the teaching of *Fortier et al.* is described with respect to a communication system comprising a user terminal apparatus such as a telephone with a display and one or more servers, with which the user terminal apparatus can electronically communicate. The connectivity between user terminal apparatus and the servers includes for instance public switched telephone networks, local area networks and wide area networks. A speech recognition algorithm is resident on one of the servers. Speech input recorded and digitally converted at the user terminal apparatus is supplied to the speech recognition algorithm, which returns the speech analysis synthesis, in particular a textual or text representation of the speech input, which is conventionally a name, a city, state and the like.

The principle steps of the method, which will be described in more detail in the following, include (see col. 2, lines 12 to 25 *Fortier et al.*)

- a) capturing in electronic form a word spoken by the speaker;
- b) passing the word to a speech recognition algorithm;
- c) receiving from the speech recognition algorithm at least one representation of the word;
- d) displaying for the speaker as text the at least one representation of the word to permit the speaker to select a correct representation of the word from among the at least one representation; and
- e) repeating the steps of a)-c) in an event that none of the representations of the word are verified as correct, or enabling the speaker to communicate the at least one word in another way.

Fortier et al. distinguishes between speech recognition, which returns a selection of possible text responses, i.e. several text responses due to ambiguous speaker's utterance, and speech recognition, which fails and does not return any possible text responses.

In the first case, a selection of possible text responses is for instance returned because of a hard to distinguish and hence similar pronunciation (cf. col. 6, lines 53 to 55: "Ohio" and "Iowa"), or different spellings of a word having the same pronunciation (cf. col. 7, lines 1 to 6: "Stephen", "Steven", "Stevan", and "Stevon"). In this case, *Fortier et al.* suggests presenting a selection of possible text representation responses, which are recognized as being similar to the speakers utterance. Such returns of speech recognition algorithm are also known as ambiguity. Typically speech recognition processes determine likelihood or probability quantities. A likelihood quantity represents a confidence value associated with a text representation of an analyzed speech input. Each text representation with a likelihood quantity exceeding a threshold value is returned by the speech recognition algorithm. The speaker is then enabled to select the appropriate one of the pre-selection of possible text representation responses, which pre-selection is based on the likelihood quantities of the text representation responses thereof.

In the latter case, the speech recognition fails and does not return any possible text response. The speech recognition for instance fails because of an unusual accent of the speaker or the voice input of an unusual name, which the speech recognition is not equipped to recognize (cf. col. 7, lines 25 to 28). Thus the speaker is not enabled to select any appropriate text representation response because the speech recognition does not return any proposal of a text representation. In the context of this case *Fortier et al.* suggests a backup procedure, which allows the speaker to verbally spell (col. 7, line 23) or to manually spell (col. 7, lines 39-40) the name (or word) to be inputted.

According to *Fortier et al.*, verbally spelling means that the speech recognition is provided with an alpha recognition algorithm, which is capable to recognize verbally spoken letters of an alphabet in the language of interest of the speaker (see col. 7, lines 30 to 38 and col. 11, lines 18 to 22 referring to steps 182 and 184 of Fig. 4e).

According to *Fortier et al.*, manually spelling means that keys of a keypad of the apparatus used for speech input such as a telephone is to be used by the speaker for

manually inputting the letters of the unrecognized spoken word (name). *Fortier et al.* describe exemplarily the use of an ITU-T keypad, which keys thereof have assigned different letters in dependency of the numbers of key presses. This letter assignment to number keys of an ITU-T keypad is known to be used for inputting short messages of the short message service (SMS) (see col. 7, lines 41 to 53 and col. 11, lines 30 to 35 referring to steps 186 et seq. of Fig. 4e and further referring to the description of Fig. 3 in col. 7, lines 41 et seq.).

Fortier et al. describes the above summarized teaching also in col. 9, lines 10 to 54. In this context *Fortier et al.* explicitly describes the meaning of the phrase “failed speech recognition”. It should be noted that it will be further explained below that the meaning of the phrase “failed speech recognition” is of particular relevance in view of the claimed subject of the present application.

As described with reference to steps 102 and 104 of Fig. 4b, a voice input is provided to the speech recognition, which response may comprise either an empty response or at least one text representation of the voice input.

Successful speech recognition according to *Fortier et al.*:

If only one representation is returned in the response (response “N” to decision step 110 if Fig 4b), the speech recognition process is considered to be successful and the returned text representation is displayed (step 112) to the speaker for verification (steps 114 and 116).

If more than one text representations are returned, the response of the speech recognition process is considered by *Fortier et al.* as successful (steps 126 and 128 of Fig. 4c) only if the number of returned representations does not exceed a predetermined limit. If the number of returned representation does not exceed the predetermined limit, the returned representations are displayed for selection by the speaker, who controls the displaying and selection with the help of key presses (steps 130 to 140). The speaker is enabled and requested to browse through the displayed representations and to select the correct one thereof. If the speaker does not select one of the displayed representations, the speaker can exit the operation sequence and the speech recognition is initialized (step 144) for restarting the speech recognition process.

Speech recognition is considered to have failed according to *Fortier et al.*:

If the predetermined limit is exceeded by the number of returned text representation, the speech recognition **is considered to have failed** (cf. col. 9, lines 40 to 42 and lines 46 to 49; step 122: “**recognition failed message**”). Upon the determination that the speech recognition is considered to have failed, the above explained alternative entry options, i.e. verbally or manually spelling, is presented to the speaker (cf. col. 9, lines 43 to 44 and lines 41 to 46 referring to Fig. 4e).

In case of the recognition **is considered to have failed**, i.e. the speech recognition returns an empty response or the number of returned representations exceeds the predetermined limit, the above explained alternative entry options, i.e. verbally or manually spelling, is presented to the speaker (cf. col. 9, lines 14 to 19 and lines 41 to 46 referring to Fig. 4e).

The teaching of *Fortier et al.* can be summarized in that the speech recognition returns one text representation or a number of text representations if the number does not exceed a predetermined limit. The latter case occurs due to ambiguity and the returned text representations are presented to the speaker for selecting one text representation thereof.

The speech recognition **is considered to have failed**, if the speech recognition returns a number of text representations exceeding the predetermined limit or if the speech recognition returns an empty response. If the speech recognition **is considered to have failed** the **alternative entry options** verbally or manually spelling are presented to the speaker. In particular, a list of text representations is not provided to the speaker for selecting thereof.

Gerson (US 6,868,385 B1) has been previously cited by the Examiner as the closest prior art and has been discussed in detail in the response to the last preceding Office Action. In the present Final Office Action, *Gerson* is referred to by the Examiner to give reasons for the assertion that speech recognition is not only used for controlling retrieval of contacts and dialing of telephone numbers thereof but also useable with further applications executable on a communication device. In this context, the

Examiner refers to col. 9, line 66 to col. 10, line 4, where *Gerson* describes that the recognized utterances (typically words) may be passed to various applications for further processing.

Applicant's Comments

The claimed concept of the present application differs from that *Fortier et al* in that a list-based manual back-up procedure is provided to the user in case the speech recognition fails. Failing of the speech recognition means that the speech recognition does not return any result.

This is contrary to the teaching of *Fortier et al.* summarized above. In accordance with the teaching of *Fortier et al.* several text representations are presented to the speaker and the speaker is requested to select one of the text representations thereof only in one specific case, namely in case the speech recognition has successfully returned a response including a number of text representations, which number does not exceed a predetermined limit. A number of text representations may be returned by the speech recognition due to ambiguity.

As aforementioned, in case of ambiguity, the speech recognition returns all text representations, which have been associated likelihood values exceeding a threshold value, which likelihood values represent confidence values that the utterance and/or pronunciation of the text representations correspond to that of the speech input by the speaker. The return of several text representations is a contribution to the limited recognition quality of the algorithms used by the speech recognition. The better the speech recognition algorithms, the higher the threshold value can be selected. As a consequence thereof, the speech recognition returns several text representations only in exceptional cases.

In the context of the speech recognition processing, *Fortier et al* uses the wording “to fail”, i.e. with more particularity “considered to have failed,” to described that the speech recognition returns either a number of text representations exceeding a predetermined limit or an empty response (see col. 9, line 42). In both cases, the response of the speech recognition is not considered for further processing. The phrase “considered to have failed” as used by *Fortier* means that either the response of the speech recognition processing is not considered because the number of

returned test representations exceeds the predetermined limit or the response is empty. Therefore, the speaker is not enabled to select from a list of text representations. As back-up operations, which are provided to the speaker as alternative input methods, *Fortier* suggests the use of verbally or manually spelling.

According to the present invention, the displaying of the list of a first set of data records or the displaying of a list of second data records is performed in case the speech recognition has failed. This back-up operation is performed as an alternative input method to the speech recognition, which means that the back-up operation does neither require any input from the speech recognition nor considers or processes any response of the speech recognition.

Contrary to the teaching of *Fortier*, the present invention nevertheless suggests the displaying of a list of a first set of data records in accordance with a first manual user input generated upon user-actuation of a multiple switching component or the displaying of a list of a second set of data records in accordance with a first manual user input generated upon user-actuation of a multiple switching component.

Upon a second manual user input, the one data record out of the displayed list is identified/selected by the user and a corresponding instruction is transmitted to the respective application, which further processes accordingly the identified data record.

According to an embodiment of the invention (cf. new claim 21), either the list of the first set of data records or the list of the second set of data records is displayed to the user. According to a further embodiment of the invention (cf. claim 2), the list of the first set of data records relates to contact information including telephone numbers and the list of the second set of data records relates to control information for controlling functions or further applications executed on the communication device. Neither the selective displaying of either the list of the first data record or the list of the second set of data records nor the list of the first data record relating to contact information and list of the second data record relating to further application function is described or suggested by the cited prior art. Particularly, the categorization of the list of first set of data records and the list of the second set of data records in combination with the selective display one of the lists thereof takes account of the usability of the back-up operation as suggested by the present invention.

According to yet another embodiment (cf. claim 22), the user is enabled to browse through the displayed list for selecting one of the data records thereof. This means that the user preselects the list and hence the category from which a selection is intended. According to embodiments of the invention, the categories include the aforementioned contact information or further application functions.

According to yet other embodiments (cf. claim 23 and 24), the list of the first data records includes all contact information selectable and activateable by speech recognition and the list of the second set of data records includes all further application functions selectable and activateable by speech recognition.

Recap and Further Comments

In Applicant's view, the meaning of the claimed "back-up operation" in conjunction with "failure of said speech recognition of said acoustic input" is a key point. From the description of the invention, it will be evident by virtue of the way the claim language is set forth that the back-up operation is completely independent from the speech recognition, which means that the back-up operation is also completely independent from any result returned by the speech recognition. More particularly, this implies that the back-up operation does not process any results returned by the speech recognition.

As explained in detail above, the Examiner considers the return of more than one text representations from the speech recognition, from which the speaker may select one text representation, as failure of said speech recognition of said acoustic input. In Applicant's view, this interpretation of the phrase "failure of said speech recognition of said acoustic input" is not supported by *Fortier* and the general understanding thereof, respectively.

In accordance with the understanding of *Fortier*, failure of said speech recognition of said acoustic input implies that any results thereof are not further considered. Rather, independent alternative input methods are suggested. In case of *Fortier* the verbally or manually spelling. In case of the subject matter of the present application, the selection and displaying of a list of (either) first set of data records or second set of data records.

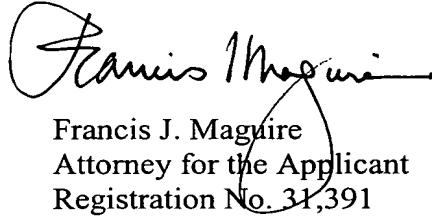
For the sake of distinguishing the subject matter of independent claim 1 from the teaching of *Fortier et al.* Applicant has amended the previously claimed "performing a back-up operation" to "performing a back-up operation alternatively to the speech recognition" in order to define that the back-up operation is independent from any results of the speech recognition.

Therefore, the Applicants respectfully request the Examiner to reconsider in light of the above and withdraw the obviousness rejection.

Enclosed herewith is an RCE Transmittal and Fee Transmittal for 12 new dependent claims accompanied by our check for \$1,410.00. If either of these Transmittals and/or fee is missing or incorrect in some way please deduct the appropriate amount from our Deposit Account No. 23-0442.

The objections and rejections of the Office Action of June 26, 2008, having been obviated by amendment or shown to be inapplicable, withdrawal thereof is requested and passage of claims 1-6 and 9-20 to issue is earnestly solicited.

Respectfully submitted,



Francis J. Maguire
Attorney for the Applicant
Registration No. 31,391

FJM/lk
WARE, FRESSOLA, VAN DER SLUYS
& ADOLPHSON LLP
755 Main Street, P.O. Box 224
Monroe, Connecticut 06468
(203) 261-1234